HOKKAIDO UNIVERSITY

Title	A Study on Machine Learning Algorithms Using Feature Interactions [an abstract of dissertation and a summary of dissertation review]
Author(s)	新, 恭兵
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第14581号
Issue Date	2021-03-25
Doc URL	https://hdl.handle.net/2115/81146
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Kyohei_Atarashi_review.pdf, 審査の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　新　恭兵

審査担当者　主　査　准教授　小山 聡
副　査　特任教授　栗原 正仁
副　査　教　授　山本 雅人
副　査　教　授　川村 秀憲
副　査　教　授　小野 哲雄

学位論文題名

A Study on Machine Learning Algorithms Using Feature Interactions
（特徴の組合せを扱う機械学習アルゴリズムに関する研究）

一般的な機械学習の枠組みにおいて、データは特徴とよばれる数値を並べた特徴ベクトルの形で与えられる。与えられた特徴をそのまま用いるのでなく、特徴の組合せを新たな特徴として用いることで、機械学習に用いるモデルの表現能力を拡大することができる。特徴の組合せを用いた機械学習は、与えられた二つのデータ間にある関係が成り立つか否かを予測するリンク予測等の問題において、高い予測精度を実現できることが従来から知られてきた。また、モデルの高い解釈性など今日の機械学習において求められている要件を満たすことが容易であることから、その重要性は一層高まっていると考えられる。一方で単純に組合せ特徴を用いると、入力データの特徴数を組合せの次数で冪乗した計算量がかかる。これを解決するためのカーネル法も提案されているが、その場合にはデータ数に対するスケーラビリティが悪化するため、大規模データでの利用が困難となる問題が存在した。

本論文は機械学習において特徴の組合せを効果的・効率的に用いるための研究をまとめたものであり、大きく二つの貢献が認められる。一つ目の貢献は、特徴の組合せを用いてリンク予測を行うための、従来の方法よりも予測精度が高く高速な手法を提案したことである。具体的には、リンク予測を行う二例間で特徴の高次の組合せを用いるカーネル関数を提案し、その効率の良い計算アルゴリズムを設計している。二つ目の貢献は、特徴の組合せを用いるカーネル関数を近似するランダム特徴写像を設計し、写像した空間で線形モデルを学習することで、特徴の組合せを用いるカーネル法のデータ数に対するスケーラビリティの問題を解決したことである。具体的には Itemset カーネルと呼ばれる特徴の組合せを用いる一般的なカーネル関数に対するランダム特徴を設計し、その近似誤差を理論的に評価するとともに、高速で省メモリに動作する計算アルゴリズムを提案している。

本論文の構成は以下のとおりである。

第 1 章では、本研究の背景と目的、本研究の貢献の概要が述べられている。

第 2 章では、線形モデルやカーネル法、Factorization Machines (FM) といった特徴の組合せを用いる主要なモデルとその最適化アルゴリズムについて述べられている。

第 3 章では、データの特徴に基づいたリンク予測において、特徴の組合せを用いる新たな手法を提案している。従来手法より高精度な予測を行うために、二例間の高次の特徴の組合せを用いる

カーネル関数を提案し、その効率の良い計算アルゴリズムを設計するとともに、カーネル関数を用いた機械学習モデルを提案している。提案手法と既存の研究の関連について整理し、さらに提案手法を取り入れた深層ニューラルネットも提案している。最後にいくつかの実データを用いた実験で提案手法の有効性を示している。

第 4 章では、組合せ特徴を用いるカーネル法のスケーラビリティを改善するランダム特徴に基づく手法を提案している。ランダム特徴とは、カーネル関数を内積の意味で近似するような高次元空間に、入力された特徴ベクトルをランダムな行列やベクトルを用いて写像する方法である。写像後の空間で線形モデルを学習することで、データ数に対するスケーラビリティを改善しつつ、カーネル関数を用いた手法に近い予測精度を実現できることが知られているが、特徴の組合せを用いるカーネルに対応するランダム特徴はこれまで知られていなかった。本論文では、特徴の組合せを扱う様々なカーネル関数の一般化である Itemset カーネルに対しランダム特徴を提案することで、特徴の組合せを用いることができ、データ数に対してスケーラビリティのある手法を実現している。また、提案したランダム特徴の近似誤差や、ランダム特徴を計算するときに用いることのできる確率分布に関する理論的な考察を述べている。さらに、より高速で省メモリに動作するランダム特徴の計算アルゴリズムと、より広いクラスのカーネル関数である重み付き Item-multiset カーネルへの拡張も行っている。最後に、提案手法が既存手法とほとんど劣らない精度でより高速に動作しすることを実データおよび人工データを用いた実験で確認している。

第 5 章では本論文のまとめと今後の展望について述べている。

これを要するに、著者は、高次の特徴の組合せを用いるカーネル関数を提案しその効率的な計算アルゴリズムを設計することで、従来の手法よりも高精度なリンク予測を可能とするとともに、特徴の組合せを用いる一般的なカーネル関数を近似するランダム特徴写像を設計することで、データ数に対するスケーラビリティの改善を実現したものであり、人工知能における機械学習の分野に対して貢献するところ大なるものがある。よって著者は、博士 (情報科学) の学位を授与される資格あるものと認める。